Method for identification of geopulses to include into the Geophysical Signal Catalogue

M.I. Gapeev[1], Yu.I. Senkevich[1], O.O. Lukovenkova[1], A.A. Solodchuk[1]

[1]Institute of Cosmophysical Research and Radio Wave Propagation FEB RAS, Mirnaya str. 7, Paratunka, Kamchatka region, Russia, 684034

Abstract. A new system approach to the identification and systematization of geophysical pulses is described. It includes stages of detection, analysis, object and structural description, pulse classification. A method based on the adaptive threshold calculation is proposed for pulse detection in geoacoustic emission signals. The authors propose to analyze detected pulses using sparse approximation methods, in particular, the method of the adaptive matching pursuit. This method allows one to decompose a pulse into basis functions of combined Gauss-Berlage dictionary with minimum spatial and temporal costs and with the required accuracy of constructed approximations. The obtained sparse representations are described under the object approach as a combination of informative features identified during the analysis, for example, the number of functions in decomposition, function parameters etc. The object description of geoacoustic emission pulses is supplemented by an original structural description based on the relations of pulse local extremums. That made it possible to reduce significantly the variety of pulse shapes for their further identification. The results obtained during the application of the described system approach to the analysis of geoacoustic signals are summarized in a Geophysical Signal Catalogue.

1. Introduction

Generally, systematization of geophysical signals is the main part of a system approach to geophysical data processing. The diversity of methods and forms of geophysical data acquisition and representation is subject to the legal basis of metrological support. When collecting data, one should meet the requirements, norms and rules of the applied measurement models according to established standardized methods. Very often, researchers face the necessity of creation of new methods and techniques for investigation of natural objects and phenomena that motivates to develop specific forms of measurements and to adjust them to the known standards. In the result, huge electronic data stores are formed for many years of observations. They are usually inaccessible for wide scientific and engineering communities owing to the absence of standards and formats for data storage and interpretation.

In view of this, the scientists of the Laboratory of Acoustic Research at the Institute of Cosmophysical Research and Radio Wave Propagation (IKIR) FEB RAS decided to unify the data collected during long-term observations (2001-2018). The initial phase of this scientific-technical work was systematization of different pulse geophysical signals. Based on the investigations which had been carried out earlier [1–10], we suggested to apply a system approach for the investigation and description of geophysical pulses and for the formation of Geophysical Signal Catalogue (hereafter Catalogue). In order to realize it, current scientific groundwork and experience in geoacoustic emission (GAE) signal investigation were applied during the first phase [8, 11–17]. The aim of the scientific-technical works realized at this stage was systematization and classification of GAE pulses on the basis of system approach to the investigation and description of the variety of forms (patterns) of these pulses.

The process of geophysical signal description may be divided into two phases: detection of signal fragments containing useful data; analysis of signal inner frequency-time structure followed by detection of characteristic features.

2. Methods and algorithms for data processing and analysis

In order to detect single pulses in geoacoustic emission signals, we developed an algorithm based on the application of adaptive threshold calculated by root-mean-square deviation (RMSD) [18]. Threshold values are calculated in disjoint windows of fixed length n by the formula

$$S_k = S_{k-1} + A \cdot (\sigma_k - \sigma_{k-1}),$$

where S_k is a threshold value in a window from k n-th to $k \cdot n+n-1$-th point, σ_i is the – RMSD calculated in a window from $(i-1) \cdot n$-th to $i \cdot n-1$-th point, A is the experimentally determined parameter. As long as the estimated threshold should depend only on background noise level, we suggest not to include the signal fragments containing pulses into the process of threshold function calculation. To determine the pulse boundaries, a signal is examined forward and backward from the point of its intersection with a threshold by a moving window with length ΔPulse onset and cutoff times are determined by the moments when all the points in a window Δ turn to be less than the defined value S_0:

$$S_0 = B \cdot \sigma_\Delta,$$

where σ_Δ is the RMSD on the given fragment of a signal, B is the experimentally defined parameter. The window length Δ is also experimentally defined.

To distinguish the pulses, we suggest making preliminary processing which consists in signal standardization and centering and analysis by sparse approximation methods.

Solution of sparse approximation problem assumes signal $s(t)$ representation in the form of a linear combination of minimum possible number of functions $g_m(t)$ from a preliminary selected basis (dictionary).

$$s(t) = \sum_{m=1}^{N} c_m g_m(t), \ \|c\|_0 \to \min, \tag{1}$$

where **c** is the decomposition coefficient vector.

Geoacoustic pulses are additive signals formed by overlapping of elementary single-frequency pulses on each other. Thus, sparse representation is an appropriate description for typical geoacoustic signals.

Approximation accuracy and sparsity depends directly on a selected function dictionary. A dictionary, composed of modulated and shifted Gaussian and Berlage functions, is appropriate for GAE signal analysis.

Gaussian pulse is defined by analytical expression

$$g(t) = A \cdot \exp(-B(t_{end}) \cdot \Delta \cdot t^2) \cdot \sin(2\pi f t),$$

where A is the amplitude chosen so that $\|g(t)\|_2 = 1$; t_{end} is the atom length; f is the frequency from 200 to 20 000 Hz (recorded frequency range); $B(t_{end})$ is the parameter B limit value calculated by the formula

$$B(t_{end}) = -\frac{4 \cdot \ln 0.05}{t_{end}^2};$$

Δ is the coefficient of parameter B variation relatively the limit value.

Berlage pulse is defined by analytical expression

$$g(t) = A \cdot t^{n(p_{\max})\Delta} \cdot \exp\left(-\frac{n(p_{\max}) \cdot \Delta}{p_{\max} \cdot t_{end}} \cdot t\right) \cdot \cos\left(2\pi f t + \frac{\pi}{2}\right),$$

where A is the amplitude chosen so that $\|g(t)\|_2 = 1$; t_{end} is atom length; p_{max} is the location of a maximum relatively atom length, $p_{max} \in [0.01, 0.4]$; f is the frequency from 200 to 20 000 Hz; $n(p_{max})$ is the n parameter limit value calculated by the formula

$$n(p_{\max}) = \frac{\ln 0.05}{\ln \frac{1}{p_{\max}} - \frac{1}{p_{\max}} + 1};$$

Δ is the coefficient of parameter n variation relatively the limit value.

Sparse approximation problem (1) is unsolvable for polynomial time. Exact solution algorithm requires complete enumeration of all possible function combinations from a dictionary, i.e. it has factorial complexity O(N!). The Matching Pursuit (MP) algorithm is one of the most frequently used algorithms for approximate solution of sparse approximation problem. The algorithm was suggested in the paper [19]. The algorithm can be written in the form of the following procedure:

$$\begin{cases} R_0(t) = s(t), \\ (m,h) = \arg\left[\max\limits_{k,j} \left\langle g_k(t-\tau_j), R_i(t) \right\rangle\right], \\ R_{i+1}(t) = R_i(t) - \left\langle g_m(t-\tau_h), R_i(t) \right\rangle \cdot g_m(t-\tau_h), \end{cases}$$

where $R(t)$ is the residual, τ is the function $g(t)$ shift relatively a signal $s(t)$.

Unfortunately, the matching pursuit algorithm has a number of significant disadvantages. Firstly, to provide decomposition sufficient accuracy, we need to apply dictionaries of larger sizes that entails power growth of algorithm execution speed. Secondly, as long as functions are selected from an unchangeable dictionary, the obtained decompositions differ by «rough» sampling in parameter space. In order to solve these problems, we suggested improving the classical algorithm so that we could make decomposition of the required accuracy on a dictionary of a limited size. As long as function parameters having the highest scalar product with a signal are defined at each iteration of the algorithm, the matching pursuit iteration may be described in the form of a problem of search for function maximum of many variables

$$F(\tau, \mathbf{p}) = \left\langle s(t), g(t-\tau, \mathbf{p}) \right\rangle \to \max_{\mathbf{p}}.$$

The main idea of the suggested improvements is the application of optimization methods to define the parameters **p** of the function having the maximum scalar product with a signal. The developed algorithm was called «Adaptive Matching Pursuit» (AMP) [15; 20; 21].

Thus, in the result of the analysis within the framework of the object approach, each signal may be an object with the following attributes, reflecting signal characteristic features:

- Length (points).
- Date and time of signal recording (with the accuracy to ms).
- Number of functions contained in N pulse.
- Resulting error (in %) calculated by the formula $ERR_N = \|\mathbf{R}_N\| / \|s(t)\| \times 100$ %.
- The following items are given for each function:
  a. type (Berlage or Gaussian);
  b. shift τ (points; if the value is negative a function is shifted to the left relatively a pulse);
  c. basic length (maximum possible function length, length and maximum location are calculated relatively this value);
  d. length t_{end} (% relatively the basic length (c));
  e. maximum location for Gerlage functions p_{max} (% relatively the length (d));
  f. filling harmonic frequency f (Hz);
  g. variability coefficient Δ (affects the envelope function steepness: the higher the value is the steeper the envelope is).

To automate the process of geoacoustic pulse characteristic feature detection, we apply «MPComplex» software. It includes four subsystems: dictionary generation subsystem, DictioanyConstructor; signal analysis subsystem, MPAnalyzer; visualization subsystem MPVisualisator and SIGView [20].

We decided to supplement the object and frequency-time description of GAE pulses with original description of amplitude-phase structure based on pulse local extremum relations. Such an approach allows us to represent a signal within the framework of a closed set without threshold values of amplitude quantization level and sampling period arising during the process of analogue signal digitization. Measurement accuracy improvement requires expansion of the indicated digitization characteristics and entails combinatorially-dependent variety of signal acceptable values. It is possible to avoid such a situation if a signal is transformed into a sequence of its local extremums and intervals between them. Such a transformation does not seem to be critical if we may not take into account the signal values between its neighbor local extremums that is quite justified in measurement practice of many natural signals. The transformation essence is illustrated in figure 1.

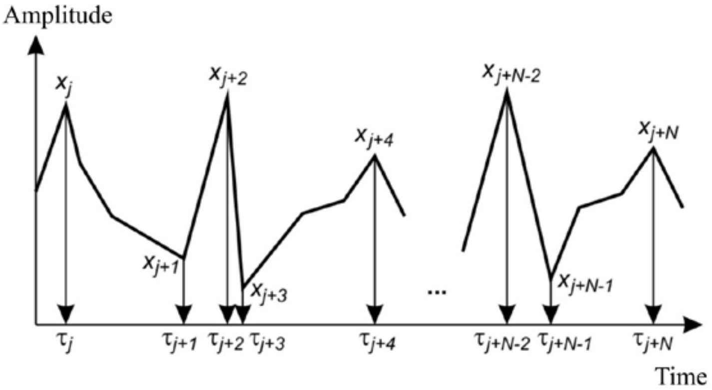

Figure 1. Example of signal digitization by the values of local extremums and intervals between them.

We have an initial set of successive values of local extremums $\{x_i\}$, where i=1,N. We supplement this set by one more set with successive values of time intervals $\{\tau_i\}$, where i=1,N–1, between local extremums. We calculate the relations for each extremum as follows:

$$r_{i,i+m} = \begin{cases} 1, x_i > x_{i+m} \\ 0, x_i \le x_{i+m} \end{cases}, \ \omega_{i,i+m} = \begin{cases} 1, \tau_i > \tau_{i+m}, \\ 0, \tau_i \le \tau_{i+m} \end{cases}, \text{ where } m = 1...M(M \le N-i), \tag{2}$$

where $r_{i,i+m}$ is the result of logical comparison of i-th and $i+m$-th values of extremum amplitudes; $\tau_{i,i+m}$ is the result of logical comparison of i-th and $i+m$-th interval values between the extremums. We put in order the series of such relations in the form of square matrixes for M-relations $r_{i,j}$ and $\omega_{i,j}$. The obtained matrixes have diagonal symmetry owing to the algebraic property of inequality symmetry (if $a>b$, then $b<a$), and are redundant in this sense, thus, we will apply only halves of each of the matrixes.

$$\mathbf{R}_i = \begin{pmatrix} r_{i,i} & r_{i,i+1} & \cdots & r_{i,i+(M-2)} & r_{i,i+(M-1)} \\ r_{i+1,i} & r_{i+1,i+1} & & & r_{i+1,i+(M-1)} \\ \vdots & & \ddots & & \vdots \\ r_{i+(M-2),i} & & & r_{i+(M-2),i+(M-2)} & r_{i+(M-2),i+(M-1)} \\ r_{i+(M-1),i} & r_{i+(M-1),i+1} & \cdots & r_{i+(M-1),i+(M-2)} & r_{i+(M-1),i+(M-1)} \end{pmatrix}, \tag{3}$$

$$\mathbf{W}_i = \begin{pmatrix} \omega_{i,j} & \omega_{i,i+1} & \cdots & \omega_{i,i+(M-2)} & \omega_{i,i+(M-1)} \\ \omega_{i+1,i} & \omega_{i+1,i+1} & & & \omega_{i+1,i+(M-1)} \\ \vdots & & \ddots & & \vdots \\ \omega_{i+(M-2),i} & & & \omega_{i+(M-2),i+(M-2)} & \omega_{i+(M-2),i+(M-1)} \\ \omega_{i+(M-1),i} & \omega_{i+(M-1),i+1} & \cdots & \omega_{i+(M-1),i+(M-2)} & \omega_{i+(M-1),i+(M-1)} \end{pmatrix}. \tag{4}$$

Matrixes (3) and (4) represent a code of the chosen (i-th) extremum in a signal characterizing its amplitude and time location in relation to the following M-extremums. As a consequence of application of relation (2) rule, transformation of a signal fragment from M-extremums into matrixes

(3) and (4) has invariant property to shift operations as well as initial signal amplitude and time transposition. The obtained important property of insensitivity to matrix shift follows from inequality main property: if $a>b$, then $a+c>b+c$ for any c for the operation of signal time shift, and if $a>b$ and $c>0$, then $ac>bc$ for signal transposition operation. In the result, some graphic invariant of signal form (pattern) can be related to each obtained pair of matrixes (3) and (4). Such a description gives the information on signal structure in relations of its extremums.

When applying the transformation, we can describe any fragment of a signal. An example of such structural description is illustrated in figure 2.

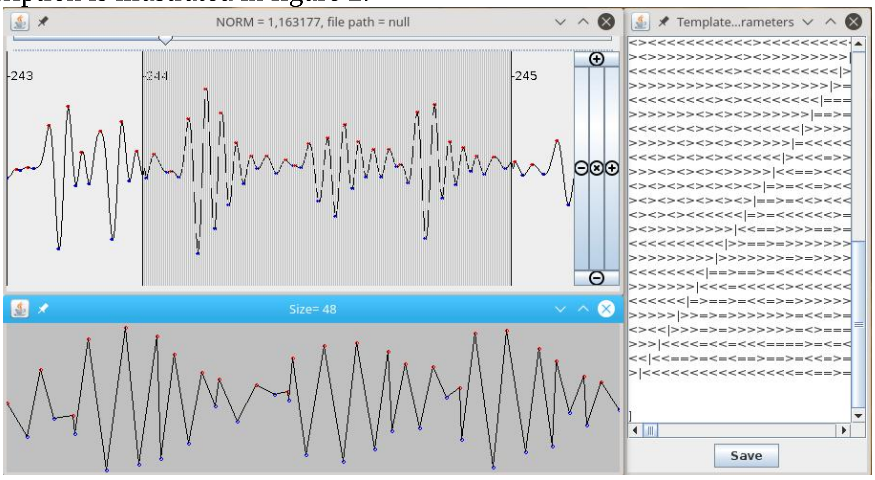

Figure 2. Examples of structural description of a signal fragment. At the top on the left is an episode of a digitized signal with a selected fragment; on the right is a matrix of selected fragment relations in symbols «greater than», «less than» or «equal to»; at the bottom is a selected fragment pattern constructed on the basis of relational matrix.

3. Results

A new approach has been proposed to describe the geoacoustic emission pulses structure and to create a Geophysical Signals Catalogue on its basis. During the process of realization of the described approach to describe GAE measurement results, we obtained unique data on the diversity, specification and characteristics of geoacoustic signals in near-surface rocks. To automate the description process of a pulse set included into the Catalogue, a software application «Registry» was written. Currently, the number of pulses included in the Catalogue is 40284.

The developed method of geoacoustic pulse detection and identification uses indicators which are invariant to the amplitude and time transposition of the detected pulse forms. Thus, this method is fundamentally different from the techniques previously applied to study geoacoustic emission signals.

4. Conclusions

- Based on the developed object description of pulses applying decomposition in a combined dictionary of Gaussian-Berlage functions, analyzed signals were decomposed into basic functions with high accuracy.
- Owing to the developed format of pulse characteristics representation in the form of a unified pattern, we demonstrated the possibility of graphic and self-explanatory description of GAE pulse set, their automatic search and sorting based on a defined pattern field set.
- By the means of structural description of selected pulses in the form of specific matrix-pattern, we demonstrated clearly the countability of geoacoustic pulse endless variety.
- In the result of geoacoustic signal processing applying the developed pattern, we created an object base for their further analytical examination in classes based on semantic information binding and detection of group properties.
- Finally, on the basis of object systematization by characteristic features and application of the object approach to data representation, Geophysical Signal Catalogue has been formed.

5. References

[1] Larionov, I.A. Geoacoustic emission response to earthquake preparation at various observation points / I.A. Larionov, A.O. Shcherbina, M.A. Mishchenko // Vestn. KRAUNTS. Earth Science Series. – 2005. – Vol. 2(6). – P. 108-115. (in Russian).

[2] Kuznetsov, Yu.V. Correlation method for detecting information components of secondary electromagnetic radiation of technical means / Yu.V. Kuznetsov, A.B. Baev, M.A. Behtin, A.A. Sergeev // Tekhnologiya EMS. – 2009. – Vol. 2(29). – P. 2-13. (in Russian).

[3] Christine, E. Estimation of soil properties using waveforms of seismic surface waves // Patent US8892410B2. – 18.11.2014.

[4] Rabinovich, E.V. Model of seismic impulse arising from hydraulic fracturing / E.V. Rabinovich, K.S. Ganchin, I.M. Pupyshev, G.S. Shefel // Mathematical structures and modeling. – 2014. – Vol. 4(32). – P. 105-111. DOI: 10.7868/S0016794015040136.

[5] Pulinets, S.A. Physical bases of the generation of short-term earthquake precursors: A complex model of ionization-induced geophysical processes in the lithosphere-atmosphere-ionosphere-magnetosphere system / S.A. Pulinets, D.V. Davidenko, D.P. Ouzounov, A.V. Karelin // Geomagnetism and Aeronomy. – 2015. – Vol. 55(4). – P. 521-538.

[6] Filonenko, S. Model of acoustic emission signal at the prevailing mechanism of composite material mechanical destruction / S. Filonenko, T. Nimchenko, A. Kosmach // Aviation. – 2010. – Vol. 14(4). – P. 95-103. DOI: 10.3846/aviation.2010.15.

[7] Mishchenko, M. Statistics of occurrence of pre-seismic anomalies in geoacoustic emission and in atmospheric field // E3S Web of Conferences. – 2016. – Vol. 11. – P. 00015. DOI: 10.1051/e3sconf/20161100015.

[8] Tristanov, A.B. A new approach to study of geoacoustic emission signals / A.B. Tristanov, Yu.V. Marapulets, O.O. Lukovenkova, A.A. Kim // Pattern Recognition and Image Analysis. – 2016. – Vol. 26(1). – P. 34-44. DOI: 10.1134/S1054661816010259.

[9] Riabova, S.A. Seismic background disturbance due to geomagnetic impulses / S.A. Riabova, A.A. Spivak // Geophysical research. – 2017. – Vol. 18(2). – P. 65-76. DOI: 10.21455/gr2017.2-4.

[10] Senkevich, Yu.I. Information approach to the analysis of acoustic and electromagnetic signals / Yu.I. Senkevich, V.A. Duke, M.A. Mishchenko, A.A. Solodchuk // E3S Web of Conferences. – 2017. – Vol. 20. – P. 02012. DOI: 10.1051/e3sconf/20172002012.

[11] Kuptsov, A.V. Geoacoustic emission during the precursory periods of Kamchatka earthquakes / A.V. Kuptsov, I.A. Larionov, B.M. Shevtsov // Volcanology and seismology. – 2005. – Vol. 5. – P. 45-49. (in Russian).

[12] Gregori, G.P. «Storms of crustal stress» and AE earthquake precursors / G.P. Gregori, M. Poscolieri, G. Paparo, S. De Simone, C. Rafanelli, G. Ventrice // Natural Hazards and Earth System Sciences. – 2010. – Vol. 10(2). – P. 319-337. DOI: 10.5194/nhess-10-319-2010.

[13] Mishchenko, M.A. Statistical analysis of disturbances of geoacoustic emissions preceding strong earthquakes in Kamchatka // Vestn. KRAUNTS. Fiz.-mat. science. – 2011. – Vol. 1(2). – P. 56-64. (in Russian).

[14] Marapulets, Yu.V. Mesoscale acoustic emission / Yu.V. Marapulets, B.M. Shevtsov // Vladivostok: «Dalnauka» Publisher. – 2012. – 126 p. (in Russian).

[15] Afanas'eva, A.A. Application of sparse approximation and clustering methods for describing the structure of acoustic emission time series / A.A. Afanas'eva, O.O. Lukovenkova, Yu.V. Marapulets, A.B. Tristanov // Digital signal processing. – 2013. – Vol. 2. – P. 30-34. (in Russian).

[16] Marapulets, Yu. Changes of geoacoustic emission directivity at «Mikizha» site associated with earthquakes in Kamchatka / Yu. Marapulets, A. Solodchuk, A. Shcherbina // E3S Web of Conferences. – 2016. – Vol. 11. – P. 00014. DOI: 10.1051/e3sconf/20161100014.

[17] Marapulets, Yu.V. Methods of recording and time-frequency analysis of geoacoustic emission signals / O.O. Lukovenkova, A.B. Tristanov, A.A. Kim // Vladivostok: «Dalnauka» Publisher. – 2017. – 148 p. (in Russian).

[18] Kim, A.A. Parallel computations for realtime implementation of adaptive sparse approximation methods / A.A. Kim, O.O. Lukovenkova, Yu.V. Marapulets, A.B. Tristanov // XX IEEE International Conference on Soft Computing and Measurements (SCM), 2017. DOI: 10.1109/SCM.2017.7970567.

[19] Mallat, S. Matching pursuits with time-frequency dictionaries / S. Mallat, Z. Zhang // IEEE Transactions on Signal Processing. – 1993. – Vol. 41(12). – P. 3397-3415. DOI: 10.1109/78.258082.

[20] Lukovenkova, O.O. A program for visualizing the results of sparse approximation of geoacoustic emission signals MPComplexver 1.0/ O.O. Lukovenkova, A.B. Tristanov // Certificate of state registration program No. 2015617250. – 29.09.2015.

[21] Tristanov, A.B. The time series model for temporal object analysis / A.B. Tristanov, O.O. Lukovenkova, R.K. Polyakov // Marine intellectual technologies. – 2016. – Vol. 4 (34). – P. 46-50. (in Russian).

Acknowledgments

The research was supported by Russian Science Foundation, project No. 18-11-00087.